Use of a Continuous Probability Scale to Display Visual Field Damage

Michael Wall, MD; Chris A. Johnson, PhD; Randy H. Kardon, MD, PhD; David P. Crabb, PhD

Objective: To derive and display the percentile score (1st-99th percentile) at each perimetric location of a standard automated perimetry test to determine whether this technique will uncover patterns of loss not visible in standard (StatPac) probability maps.

Methods: We computed continuous scale probability plots of data collected from testing 305 visually healthy participants with standard automated perimetry (24-2 Swedish interactive thresholding algorithm). The age-corrected thresholds were sorted by sensitivity at each visual field location. Percentiles were derived in single increments from the 1st to the 99th. We displayed the percentiles as a color scale and then interpreted visual field plots from healthy control subjects and patients with visual system disorders.

Results: Added information was achieved for identifying patterns of visual loss by using the 5th- to 20th-percentile range in conjunction with the lower range below the 5th percentile that is typically used. The extent of contiguous regional defects appeared larger using this method. Healthy control subjects often have threshold results within the 5th- to 20th-percentile range, but these test locations usually appeared randomly spaced rather than in contiguous patterns commonly seen in patients at the border of visual field defects.

Conclusion: Continuous scale probability plots are a potentially useful adjunct for interpretation of perimetry results.


Automated perimetry was introduced into ophthalmology in the late 1970s. Until the late 1980s, when Schwartz and Nagin and Heijl and Asman introduced the concept of perimetric probability maps, clinicians struggled to decide what constituted a perimetric defect. These empirically generated probability plots showed the likelihood that an observed threshold level belonged to an age-corrected normal population. Sensitivity values in the lower 5%, 2%, 1%, and 0.5% levels were identified on a printout. This analysis technique was embraced by clinicians and has become a clinical standard used to interpret visual field results point by point.

Why was the 5% level ($P < .05$) chosen as a cutoff for this analysis? The 5% level is the standard level of significance used to justify a claim of a statistically significant effect. It is a cutoff level commonly used to filter signal from noise when testing the null hypothesis, and the 5% represents the probability that a chance finding has occurred. This level, a standard limit in science and medicine, is thought to have its origins in a series of tables published by the statistician Fisher in 1925. Fisher stated,

Personally, the writer prefers to set a low standard of significance at the 5 per cent point, and ignore entirely all results which fail to reach this level. A scientific fact should be regarded as experimentally established only if a properly designed experiment rarely fails to give this level of significance.

The use of statistical tables with fixed probability values clearly propagated this concept.

However, a more pragmatic view of the cutoff for statistical significance is that it depends on how much risk one is willing to accept in stating that a sample is from a probability distribution when it may come from some other source. The risk one is willing to accept should depend on the situation and the consequences of being mistaken, a familiar concept in clinical decision making. Simply put, reliance on a fixed level of significance is more a result...
of statistical tables from the past; an effect should not necessarily be declared unimportant if the probability of it occurring by chance is, say, 6%.

Given that in perimetry we are more concerned about the pattern of abnormal test locations rather than whether a single test location meets the standard for clinical significance, we propose viewing probability maps on a continuous probability scale. Such an approach may allow a clinician to better judge the true extent of an abnormally depressed threshold area of visual field that is clinically significant. For example, in a patient with intraocular pressures in the mid 20s (in milligrams of mercury) with an optic disc showing an inferior notch, one should ideally let the clinician decide what conclusion to draw if there are 3 test locations occurring as a superior nasal step with each at the $P=.07$ level of significance. With current perimetry software fixed at 5% cutoff values, these values are not identified and would likely be missed. Therefore, the aim of this work, as a proof of principle, was to derive and display the percentile score (1st-99th percentile) at each perimetric location of a standard automated perimetry test to determine in a qualitative fashion whether this technique will uncover patterns of loss not visible in standard (StatPac) probability maps (Zeiss Humphrey Systems, Dublin, California) (half to fifth percentile).

Therefore, the aim of this work, as a proof of principle, was to derive and display the percentile score (1st-99th percentile) at each perimetric location of a standard automated perimetry test to determine in a qualitative fashion whether this technique will uncover patterns of loss not visible in standard (StatPac) probability maps (Zeiss Humphrey Systems, Dublin, California) (half to fifth percentile).

### SUBJECTS

The visual testing protocols were approved by the University of Iowa Investigational Review Board and the Legacy Health System (Devers Eye Institute, Portland, Oregon) Investigational Review Board. The tenets of the Declaration of Helsinki were followed. All of the subjects gave informed consent to participate in the study. The 190 healthy control subjects from Iowa were mostly paid volunteers who answered advertisements or telephone calls inviting them to participate in research; all of the visual field examinations of the healthy subjects were included unless they did not meet the manufacturer’s criteria for reliability. The mean (SD) age of the healthy control subjects was 46.6 (15.6) years (range, 10-92 years). There were 132 right eyes and 173 left eyes. The healthy control subjects from Portland were composed of 115 participants (70 women and 45 men) with normal visual fields and ranging in age from 21.6 to 92.4 years (mean [SD] age, 49.0 [13.6] years) who participated in this study. Self-reported race for the Portland subjects either had testing in 1 eye chosen at random or had both eyes tested with 1 eye chosen at random for the study. For this work, we used this table to evaluate probability plots of a series of healthy control subjects and patients with optic neuropathy, including patients with glaucoma.

Finally, the percentile levels were plotted on a color scale for each visual field and colorized with a stoplight metaphor (Figure 1A) using the statistical programming language R version 2.0.1 (The R Foundation for Statistical Computing, Vienna, Austria). The scale uses shades of green to show percentiles above the 20th percentile, yellow and orange for the 6th to 20th percentiles, and red, gray, and black for the 5th percentile or lower. A gradient of red to black denotes $P<.05$. We give the actual percentile value as an integer within the corresponding colored square (Figure 2). A noninterpolated gray scale shows the raw, non–age-corrected threshold values (Figure 1B).

A definition of abnormality can be developed with this type of analysis. We provide the Table as a beginning framework. This shows the probability of a given number of abnormal test locations at a specified percentile value occurring in a superior or inferior hemifield by chance alone up to the 20th percentile. It is important to state that this assumes that points in the visual field act independently, which is not the case. The sensitivity of locations in the visual field is spatially dependent and intrinsically related (test locations are correlated). Nevertheless, we used this table to evaluate probability plots of a sample (n=60) of the healthy control subjects. For this work, we declared a field to be overall defective if it had more than a certain number of locations below the given percentile. For example, a hemifield would be considered overall defective if it had more than 3 points flagged at the 5% level, more than 5 points defective at the 10% level, or more than 7 points at the 15% level.

### VISUAL FIELD TESTING

All of the subjects underwent testing with size 3 stimuli using the standard 24-2 Swedish interactive thresholding algorithm. We followed the manufacturer’s recommendations and used a corrective lens when necessary. Care was taken to prevent lens rim artifact. The healthy control subjects either had testing in 1 eye chosen at random or had both eyes tested with 1 eye chosen at random for the study.

### PROBABILITY PLOTS

We computed empirical probability plots of data collected from testing 305 healthy participants. We imported the data to a spreadsheet after conversion by PeriData (PeriData Software GmbH, Heurth, Germany). We then found the effect of age by regressing observed threshold on age; this was performed for each test location. We adjusted all of the threshold values to an equivalent of age 45 years. The 305 values at each test location were ranked from highest to lowest in the spreadsheet. The 1st- to 99th-percentile reference levels were empirically determined. We computed age-corrected threshold percentile levels for each visual field location for a series of healthy control subjects and patients with optic neuropathy, including patients with glaucoma.

Finally, the percentile levels were plotted on a color scale for each visual field and colorized with a stoplight metaphor (Figure 1A) using the statistical programming language R version 2.0.1 (The R Foundation for Statistical Computing, Vienna, Austria). The scale uses shades of green to show percentiles above the 20th percentile, yellow and orange for the 6th to 20th percentiles, and red, gray, and black for the 5th percentile or lower. A gradient of red to black denotes $P<.05$. We give the actual percentile value as an integer within the corresponding colored square (Figure 2). A noninterpolated gray scale shows the raw, non–age-corrected threshold values (Figure 1B).

A definition of abnormality can be developed with this type of analysis. We provide the Table as a beginning framework. This shows the probability of a given number of abnormal test locations at a specified percentile value occurring in a superior or inferior hemifield by chance alone up to the 20th percentile. It is important to state that this assumes that points in the visual field act independently, which is not the case. The sensitivity of locations in the visual field is spatially dependent and intrinsically related (test locations are correlated). Nevertheless, we used this table to evaluate probability plots of a sample (n=60) of the healthy control subjects. For this work, we declared a field to be overall defective if it had more than a certain number of locations below the given percentile. For example, a hemifield would be considered overall defective if it had more than 3 points flagged at the 5% level, more than 5 points defective at the 10% level, or more than 7 points at the 15% level.

©2009 American Medical Association. All rights reserved.
The mean (SD) sensitivity for the 305 subjects was 30.6 (2.0) dB. For individual test locations, the change in percentile as a function of age-corrected threshold sensitivity shows how, except for the tails of the distribution, small changes in sensitivity give large changes in percentile value, indicating a rather narrow distribution of threshold values for healthy subjects. For example, between the tails of the percentile range, a change in 1 dB commonly results in a change of 5 to 10 percentile units.

Figure 2 shows an example of a patient with ocular hypertension. A cluster of 5 test locations at the 17th-percentile level and below occurs in the inferior nasal area. Only 1 abnormal test location is identified by standard empirical probability plots (plot directly below the percentile graph). While this appears to be a minor detail, standard definitions of glaucoma require abnormal perimetry. Consequently, if this cluster of test locations below the 18th percentile in this patient with ocular hypertension is considered abnormal, the patient would meet criteria for glaucoma.

Figure 3 shows a patient with a meningioma causing compression of the optic chiasm. In the upper image, the visual field results before radiation therapy are shown. Three years later following treatment, the patient has a large cluster of low normal sensitivities that respects the vertical midline; analysis using the 5th percentile as a cutoff shows only 2 abnormal test locations, neither of which are along the vertical meridian. This example shows how using the pattern of test locations of the 6th- to 20th-percentile levels can show a hemianopic defect that is hidden to standard probability plot analysis.

Figure 4 shows a time series of 4 results during 6 years from a patient with glaucoma. The development of visual loss in areas that were previously normal in the 6th- to 20th-percentile range suggests some predictive value of these percentile values.

We analyzed the initial and repeat results of 60 independent visually healthy control subjects using the criteria from the Table. We found a large number of these subjects (7 of 60 subjects) to be declared overall defective. However, the defects were reproducible in only 3 of the 60 subjects (the defects were in the same place in one case, adjacent in another, and separate in the third). The results of the first 10 visually healthy subjects are shown as a representative sample in Figure 5 and Figure 6. Test 1 is shown for each visually healthy participant, followed by a repeat test within 2 months (test 2). Note the frequent presence of randomly spaced abnormal test locations in the 6th- to 20th-percentile range.

The interested reader can find more examples of continuous scale probability plots in healthy control subjects and patients with glaucoma on our Web site, http://webeye.ophth.uiowa.edu/vip/.
Present if only the 2 test locations reaching the right superior quadrant defect. Two years later, the patient has a large cluster of low normal sensitivities that respects the vertical midline; this hemianopia is not

The defect at each location is considered to be the result of a Bernoulli trial with a result being defective or not. We then consider the probability of the defect occurring to be a random variable from the binomial distribution and each location being an independent trial. For any percentile, we can determine the expected number of defects in a series of trials. If we assume independence, then we can say that a cluster of 3 or more points being defective at the \( P = .05 \) level is approximately equivalent to a cluster of 5 or more defective points at the \( P = .10 \) level. This can only be considered an illustration as visual field points are not spatially independent. The first occurrence of \( P < .05 \) for increasing number of abnormal test locations appears in white for each percentile.

To date, these probability values are displayed for only fixed levels of probability. Our results suggest that more sophisticated statistical approaches need to be developed to delineate appropriate cutoffs. For now, interpretation is much like

Table. Probability of a Given Number of Abnormal Test Locations at a Specified Percentile Occurring in a Superior or Inferior Hemifield by Chance Alone if the Measured Value at Each Location Is Considered to Be an Independent Event

<table>
<thead>
<tr>
<th>Abnormal Test Locations, No.</th>
<th>1st</th>
<th>2nd</th>
<th>5th</th>
<th>7th</th>
<th>10th</th>
<th>12th</th>
<th>15th</th>
<th>20th</th>
</tr>
</thead>
<tbody>
<tr>
<td>&gt;0</td>
<td>.23</td>
<td>.41</td>
<td>.74</td>
<td>.85</td>
<td>.94</td>
<td>.96</td>
<td>.98</td>
<td>&gt; .99</td>
</tr>
<tr>
<td>&gt;1</td>
<td>.03</td>
<td>.10</td>
<td>.38</td>
<td>.55</td>
<td>.75</td>
<td>.84</td>
<td>.92</td>
<td>.98</td>
</tr>
<tr>
<td>&gt;2</td>
<td>.002</td>
<td>.02</td>
<td>.14</td>
<td>.27</td>
<td>.49</td>
<td>.62</td>
<td>.77</td>
<td>.92</td>
</tr>
<tr>
<td>&gt;3</td>
<td>&lt; .001</td>
<td>.002</td>
<td>.04</td>
<td>.10</td>
<td>.26</td>
<td>.38</td>
<td>.56</td>
<td>.79</td>
</tr>
<tr>
<td>&gt;4</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>.009</td>
<td>.03</td>
<td>.11</td>
<td>.20</td>
<td>.35</td>
<td>.62</td>
</tr>
<tr>
<td>&gt;5</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>.002</td>
<td>.008</td>
<td>.008</td>
<td>.018</td>
<td>.42</td>
</tr>
<tr>
<td>&gt;6</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>.001</td>
<td>.003</td>
<td>.009</td>
<td>.03</td>
<td>.13</td>
</tr>
<tr>
<td>&gt;7</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>.001</td>
<td>.001</td>
<td>.002</td>
<td>.01</td>
<td>.06</td>
</tr>
<tr>
<td>&gt;8</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>.001</td>
<td>.001</td>
<td>.003</td>
<td>.02</td>
</tr>
<tr>
<td>&gt;9</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>.001</td>
<td>.003</td>
<td>.008</td>
</tr>
<tr>
<td>&gt;10</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>&lt; .001</td>
<td>.001</td>
<td>.008</td>
</tr>
</tbody>
</table>

Percentile, \( P \) Value

\( \text{Value} < .001 \), < .01, < .05, < .10, < .25, < .50, < .75, < .95

The defect at each location is considered to be the result of a Bernoulli trial with a result being defective or not. We then consider the probability of the defect occurring to be a random variable from the binomial distribution and each location being an independent trial. For any percentile, we can determine the expected number of defects in a series of trials. If we assume independence, then we can say that a cluster of 3 or more points being defective at the \( P = .05 \) level is approximately equivalent to a cluster of 5 or more defective points at the \( P = .10 \) level. This can only be considered an illustration as visual field points are not spatially independent. The first occurrence of \( P < .05 \) for increasing number of abnormal test locations appears in white for each percentile.

Figure 3. A patient with chiasmal compression due to meningioma (see inset magnetic resonance image). The visual field results before radiation therapy show a right superior quadrant defect. Two years later, the patient has a large cluster of low normal sensitivities that respects the vertical midline; this hemianopia is not present if only the 2 test locations reaching the \( P < .05 \) level are used. See Figure 1 for scale values.

Pretreatment perimetry

Posttreatment improvement

have been the common standard across automated perimetry types and are the cornerstone to clinical interpretation and analysis of perimetry results in clinical trials. To date, these probability values are displayed for only fixed levels of probability. Our results suggest that more information can be acquired by viewing these on a continuous probability scale. However, this analysis assumes that each test location acts as an independent event. It is well known that adjacent test locations in the nerve fiber bundle and hemianopic regions are correlated. The cases highlighted in this work indicate that contiguous test locations exhibiting thresholds better than the 5th percentile may show diseaselike patterns in cases of mild loss and appear to have utility in interpretation of results.

Continuously scaled age-corrected probability plots use all of the information available instead of a cutoff at the 5th percentile. Should a group of test locations at the 6th to 10th percentiles showing a clear pattern of a nerve fiber bundle or a hemianopic pattern that respects the vertical midline be accepted as normal? Our results suggest that they should not. However, more sophisticated statistical approaches need to be developed to delineate appropriate cutoffs. For now, interpretation is much like
Goldmann kinetic perimetry, where with mild loss the pattern of the defect is the most important feature. Statistical measures to detect patterns in perimetry are early in their development but are likely to become more useful as they are refined.

Another problem in automated perimetry is comparison of different tests and different methods. For example, most conventional automated perimetry testing in the first 2 decades of use was done using a staircase procedure with full threshold testing. The past decade has seen the adoption of the Bayesian-based Swedish interactive thresholding algorithm strategy. As demonstrated in our examples, use of continuous scale probability plots of age-corrected thresholds may have utility in comparing results between different perimetric platforms. However, this technique needs more evaluation as we do not believe there is sufficient evidence for its application for longitudinal analysis at this time.

Like all perimetric analysis methods, continuous scale probability plots have some weaknesses. The most obvious can be seen with the examples of healthy subjects. We found that 12% of healthy control subjects had visual field defects; in 5% of the healthy control subjects, the visual field defects were reproducible. With analysis of the standard Humphrey StatPac printouts on these 60 healthy participants, 10% had visual field defects with the same subjects (5%) being reproducible. This is in line with other studies on healthy control subjects from the general population and compares favorably with analysis of the continuous scale probability plots.11,12 Whether these apparent defects are simply a chance occurrence or whether they presage future glaucomatous or other pathologic deficits will require longitudinal follow-up of

Figure 4. Four results of a patient with glaucoma during a 6-year period. Visual loss at the 5th-percentile level later develops in areas that were previously abnormal in the 6th- to 20th-percentile range. There is progression in the inferior portion of the disc with some rim thinning and vessel displacement. See Figure 1 for scale values.
these apparently visually healthy subjects. What is of interest here is whether points can be considered spatially related. Moreover, it suggests that a more sophisticated rubric for quantifying the number of defects at less conservative levels of probability is required beyond the analysis where we assume that the points act independently. An important next analysis would attempt to incorporate one of the possible models for estimating the spatial correlations that exist between locations.\textsuperscript{13,14} Another weakness is that these plots are sensitive to small decibel changes from about the 5th to 95th percentiles. For example, a 1-dB change in sensitivity can mean a 5- to 10-point percentile change. Therefore, the strength of this technique is in detecting patterns of loss rather than in detecting change at individual test locations. Perusal of the figures shows that these patterns of abnormalities of nerve fiber bundle–like groupings of test locations are the strength of this analysis.

In summary, continuous scale total deviation empirical probability plots show clinically relevant patterns of visual loss in visual field regions that are classified as normal using a 5th-percentile cutoff. These patterns may be nerve fiber bundle–like zones or they may outline depressed areas along the vertical midline that signal central nervous system visual pathway lesions. The strengths of this concept appear to be in pattern detection and using information that is typically neglected in perimetry probability plots. Appropriate statistical methods that do not
Figure 6. The results of the second 5 visually healthy subjects are shown as a representative sample. The images to the left are from test 1, and the images to the right are from a retest (test 2). The total deviation probability plots from the Humphrey Field Analyzer (Zeiss Humphrey Systems, Dublin, California) are included for comparison. Visual loss at the 5th-percentile level later develops in areas that were previously abnormal in the 6th- to 20th-percentile range. See Figure 1 for scale values.

have to assume that points in the visual field are independent would be required to differentiate significant patterns from those occurring by chance. We conclude that this approach may provide additional graphical information that may be beneficial when interpreting perimetry results.

Submitted for Publication: August 15, 2008; final revision received September 15, 2008; accepted October 10, 2008.

Correspondence: Michael Wall, MD, Department of Neurology, University of Iowa College of Medicine, 200 Hawkins Dr, #2007 RCP, Iowa City, IA 52242-1053 (michael-wall@uiowa.edu).

Financial Disclosure: None reported.

Funding/Support: This study was supported by a VA Merit Review Grant and by an unrestricted grant to the Department of Ophthalmology, University of Iowa College of Medicine from Research to Prevent Blindness, New York, New York.

Additional Contributions: Paul Artes, PhD, provided time and expert assistance with the R programming and provided manuscript suggestions.

REFERENCES


**From the Archives of the Archives**

The patient from whom the eye was taken at the autopsy died at the age of sixty-two. Two days before death I found in both eyes marked atheroma of the retinal vessels, numerous hemorrhages, and, in the right eye, the well-known picture of retinitis proliferans. The white bands, with their elevations and depressions, lay outward from the macula, and projected forward some distance into the vitreous. The patient presented general arterio-sclerosis, and died with cerebral hemorrhages.

When the eyeball was opened an elevation was found outward from the macula, having a height of 2 mm, and a maximum breadth of 3 mm.

**MICROSCOPIC**

The retina, near the ora serrata, presents the picture of edema of the retina (Iwanoff), large cystic cavities being present. Nearer the pole of the eye the cysts become smaller, and the retina is abnormal only in the hypertrophy of Müller’s fibers. To the temporal side of the disc the cysts, here filled with colloid masses, reappear. The inner nuclear layer is here distorted and mostly destroyed, and the nerve-fiber layer has disappeared entirely. The retinal vessels have atheromatous walls, and particularly the choroidal arteries have walls so thickened as to obliterate the lumen entirely in some cases. In the elevated fold of the retina near the macula, Müller’s fibers are hypertrophied, and there are cellular deposits of various sorts. A large retinal vessel breaks up into smaller branches and capillaries along the fold, and some of these pass out of the retina and end in the vitreous among the hemorrhages, apparently being in connection with the new formations in the vitreous. About these vessels are leucocytes, spindle cells, and fibrillæ. The new formation is partially covered with a membrane, which I take to be the hyaloid, and partly adjoins directly the coagula of blood.

The left eye exhibits similar retinal changes though in less degree.

To recapitulate, the chief changes are (1) atheroma of the vessels leading to hemorrhage, (2) edema of the retina characterized by the presence of cystic cavities and hypertrophy of Müller’s fibers, and (3) a very vascular connective tissue new formation in the vitreous resting on an elevated fold in retina.